UNPUBLISHED PRELIMINARY DATA

# Some Explicit Results in Multivariate Prediction Theory*

18p

F. J. Beutler[+] [redacted]
The University of Michigan, Ann Arbor, Michigan [redacted]

AUTH-ABST 14779

Summary—For a certain class of multivariate random processes, called wide-sense Markov processes, the minimum mean-square error linear prediction and estimation problem is solved in terms of explicit formulas. These continue to apply when only irregular samples of the process are available, or the optimum filter is constrained to have finite memory. A complete characterization of these processes is obtained; a process is shown to be wide-sense Markov if and only if its covariance matrix satisfies a specified functional equation. The output of a system described by a (possibly time-varying) matrix-vector equation forced by multivariate white noise is shown to be wide-sense Markov. Prediction and estimation optima are computed for the output of such a system.

AUTHOR

GPO ____
OTS ____
HC $1.00
MF $.50



## INTRODUCTION

Increased interest has recently been focused on linear multivariate systems characterized by an input-output relationship

$$y_j(t) = \sum_{k=1}^{n} \int_{-\infty}^{t} W_{jk}(t,\tau)x_k(\tau)d\tau \qquad j=1,2,\ldots,n \;, \tag{1}$$

where the inputs are $x_1(t)$, $x_2(t),\ldots,x_n(t)$ and the outputs are similarly indexed y's. The input components $x_j(t)$ are often second-order processes, which may be nonstationary.

If the system is to be designed as an optimum predictor in the sense of minimizing the statistical expectation (indicated by prefixing E) given by $E \sum_j \{[x_j(t+\alpha)-y_j(t)]^2\}$, the $n^2$ weighting functions $W_{ij}$ are given as the solution of $n^2$ linear simultaneous integral equations of the time-varying Wiener-Hopf type. The complexity of this task is such that general solutions have not been obtained, nor is there much hope of securing solutions in the future. There are, however, a number of special assumptions under which solutions are known to exist. Although many situations of practical

*This work was supported by the National Aeronautics and Space Administration under research grant NsG-2-59.

[+] [redacted] Engineering Program, The University of Michigan.



AUG 12 1963

importance satisfy these assumptions, the optimization procedures required are so difficult as to discourage actual computation [1] [2] [3].

Perhaps the outstanding feature of the theory to be presented in this paper is its ease of application. Solutions to prediction problems are presented explicitly in closed form. The technique is easily extended to the optimum (minimum mean-square) estimator for functionals on x(t) provided by unrealizable operations of certain types. The procedure may be applied when only (possibly irregularly spaced) samples of arbitrary width are available. A further extension to a generalized mean-square error criterion is immediate.

While our theory is applicable only to a certain class of (time-varying, possibly complex) random processes, this class includes many problems of engineering interest. Moreover, we derive a criterion whereby this class of processes may be easily identified. Included, for example, is the x(t) whose components solve the sets of simultaneous differential equations

$$\dot{x}_j(t) = \sum_k a_{jk}(t)x_k(t) + \sum_k g_{jk}(t)n_k(t) \ , \tag{2}$$

where the $n_k(t)$ are uncorrelated white-noise processes (compare [3]).

## I. PRELIMINARIES

A multivariate second-order process x(t) consists of n component processes $x_1(t), x_2(t), \ldots, x_n(t)$, each of which is itself a (complex) process of finite mean square. It will be convenient throughout to regard x(t) as the matrix

$$\begin{bmatrix} x_1(t) & 0 & 0\ldots0 \\ x_2(t) & 0 & 0\ldots0 \\ \ldots & \ldots & \ldots \\ x_n(t) & 0 & 0\ldots0 \end{bmatrix} \tag{3}$$

and $x^*(t)$ as the complex conjugate transpose of x(t).

The expectation of a matrix (denoted by prefixing E) is formed of the expectation of its elements. The same convention applies to other linear operations on a matrix; for example, the derivative of $A = [a_{ij}(t)]$ is $\dot{A} = [\dot{a}_{ij}(t)]$.

The correlation of x(s) and x(t) is designated P(s,t), and is given by

$$P(s,t) \quad = \quad E[x(s)x^*(t)]\ , \tag{4}$$

which is simply the matrix whose ij element is $E[x_i(s)\overline{x_j(t)}]$. The inner product of x(s) and x(t) is then defined as the trace of its correlation, viz.

$$(x(s),x(t)) \quad = \quad \tau[P(s,t)] \quad = \quad \sum_i E[x_i(s)\overline{x_i(t)}]\ . \tag{5}$$

The inner product is also defined for random variables obtained from linear matrix operations on x(t). Thus, if y = A(s)x(s) and z = B(t)x(t), the inner product

$$(y,z) \quad = \quad (A(s)x(s),B(t)x(t)) \quad = \quad \tau\{E[A(s)x(s)x^*(t)B^*(t)]\}$$

$$\tau\{A(s)P(s,t)B^*(t)\}.$$

The inner product leads also to a norm; thus the norm $\|x(t)\|$ of x(t) is given according to the definition

$$\|y(t)\|^2 \quad = \quad (y(t),y(t)) \quad = \quad \sum_i E|y_i(t)|^2.$$

The norm of z(t) above is then seen to be $\|z(t)\|^2 = \tau[B(t)P(t,t)B^*(t)]$.

Consider now a Hilbert space H; it shall consist of all matrices A(t)x(t) for each t, providing the norm remains finite, of all finite linear combinations $\sum_k A(t_k)x(t_k)$, and of all limits of such finite combinations whose norm remains finite.[1] If it is required to estimate $z(t)\epsilon H$ [$z(t) = x(t+\alpha)$ for prediction], and if $y(t)\epsilon H$ is the actual output, the error is given by z(t)-y(t). Now the jth error component is $z_j(t)-y_j(t)$, and a logical definition of the mean-square error is the sum of the mean-squares of these components. Therefore, the mean-square error $\overline{E}^2$ is

$$\overline{E}^2 \quad = \quad \|z(t) - y(t)\|^2 \tag{8}$$

[1]The reader unfamiliar with Hilbert spaces is advised to consult [4].

[redacted]

so that the optimization problem is equivalent to minimizing an H norm.

Since $z(t)$ is specified in the problem statement, the optimization procedure involves finding that "admissible" $y(t)$ which minimizes (8). To be precise, $y(t)$ is admissible if $y(t) \epsilon M$, where M is a subspace. A subspace is defined as follows: let T be a specified subset of the real line, and consider all $x(t)$ for which $t \epsilon T$. All such $x(t) \epsilon M$, together with all finite sums $\sum_k A(t_k)x(t_k)$. Finally, all limits in norm of these sums belong to M also. If t represents the time of the present, $T = \{s|s \leq t\}$ is the set of times over which $x(t)$ is available for observation or as an input to a filter; it is the "subspace of the past."

Two random variables, say y and z, are said to be orthogonal, $y \perp z$, if $(y,z) = 0$. That orthogonality is a concept central to the optimization problem is indicated by

Theorem 1: A necessary and sufficient condition that $y \epsilon M$ minimize (8) over all elements of M is that z-y is orthogonal to every element in M, i.e., that

$$z-y \perp M \qquad \text{any} \qquad y \epsilon M.$$

Such a $y \epsilon M$ exists and is unique.

Proof: See [4], Sections 6 and 7.

The requirement (9) is equivalent to $\tau\{E[(z-y)x^*(t)A^*(t)]\} = 0$ for arbitrary $A(t)$ and all $t \epsilon T$.[2] In particular, the arbitrary nature of $A(t)$ implies that (9) is satisfied if and only if the matrix

$$E[(z-y)x^*(t)] = 0, \qquad \text{all} \qquad t \epsilon T. \tag{10}$$

Suppose now that it is desired to optimize with respect to a generalized mean-square error criterion, that is, to minimize

$$\overline{E}^2 = \|B(z-y)\|^2 = E\left\{\sum_j \left| \sum_k b_{jk}(z_k - y_k) \right|^2\right\} \tag{11}$$

over all $y \epsilon M$, with B an arbitrary matrix having entries $b_{jk}$. It may be shown (see [5], Theorem 1) that the same $y \epsilon M$ which minimizes (8) also minimizes the generalized mean-square error, (11). However, the minimizing y is not unique unless B is nonsingular.

[2]Use has been made of the fact that $[A(t)x(t)]^* = x^*(t)A^*(t)$.

## II. MULTIVARIATE WIDE-SENSE MARKOV PROCESSES

The linear minimum mean-square error estimator, sometimes called a wide-sense conditional expectation, will be denoted by $\hat{E}$. For the minimum mean-square estimate of $z \in H$, based on $x(t_1), x(t_2), \ldots, x(t_{n-1})$, one may write

$$\hat{E}[z \,|\, x(t_1), x(t_2), \ldots, x(t_{n-1})] = \sum_{1}^{n-1} A_k x(t_k) , \tag{12}$$

with the matrices $A_k$ so chosen that $E[|z - \sum_{1}^{n-1} A_k x(t_k)|^2]$ is minimized. A special case is obtained by taking $t_1 < t_2 < \ldots . < t_n$ and $z = x(t_n)$. Then the process $x(t)$ is said to be wide-sense Markov if any such set of t's yields $A_k = 0$ for $k=1,2,..,n-2$; this is equivalent to

$$\hat{E}[x(t_n) \,|\, x(t_1), x(t_2), \ldots, x(t_{n-1})] = \hat{E}[x(t_n) \,|\, x(t_{n-1})] \tag{13}$$

for all $t_1 < t_2 < \ldots . < t_n$. Moreover, it may be verified that if $x(t)$ has a continuous correlation $P(s,t)$, and A is a subset of the real line such that

$$\sup_{s \in A} s = s^* < t,$$

$$\hat{E}[x(t) \,|\, x(s), s \in A] = \hat{E}[x(t) \,|\, x(s^*)] \tag{14}$$

is an equivalent definition for the wide-sense Markov property.

We remark that $\hat{E}$ is the linear minimum-mean-square-error estimate of $x(t)$, given the random process over a subset A of the real line. In general, this estimate is extremely difficult to compute. However, the property (14) sets this estimate equal to $\hat{E}[x(t)|x(s^*)]$, which is simply

$$\hat{E}[x(t) \,|\, x(s^*)] = R(t,s^*)x(s^*). \tag{15}$$

According to (10), (15) is true if $R(t,s^*)$ satisfies

$$E\{[x(t)-R(t,s^*)x(s^*)]x^*(s^*)\} = P(t,s^*)-R(t,s^*)P(s^*,s^*) = 0. \tag{16}$$

Suppose first that $P(s^*,s^*)$ is invertible. Then it is easily verified that

$$R(t,s^*) = P(t,s^*)[P(s^*,s^*)]^{-1} \tag{17}$$

satisfies (16).

If $P(s^*,s^*)$ is singular,[3] (17) makes no sense, but it is still possible to find an $R(t,s^*)$ which satisfies (15). For simplicity of notation we omit the asterisk in this discussion and speak of the determination of $R(t,s)$. Let $\lambda_1,\lambda_2,\ldots,\lambda_n$ be the eigenvalues of $P(s,s)$, and take D as the diagonal matrix whose entries are $d_{ij} = \lambda_i\delta_{ij}$ (note that s is suppressed in this discussion). Since $P(s,s)$ is Hermitian, n orthogonal eigenvectors $\{c_{i1},c_{i2},\ldots,c_{in}\}$ correspond to the eigenvalues; we may take these to be normalized by $\sum_j |c_{ij}|^2 = 1$. The unitary matrix C is then formed from the elements $c_{ij}$.

Finally, let E be the diagonal matrix with $e_{ii} = \lambda_i$ if $\lambda_i \neq 0$, and $e_{ii} = 1$ if $\lambda_i = 0$. Then

$$R(t,s) = P(t,s)[C^*E^{-1}C], \tag{18}$$

which, it may be verified, satisfies (16) as shown in [4]. It is also clear that if $P(s,s)$ is invertible, $E^{-1} = D^{-1}$, so that $C^*E^{-1}C = [P(s,s)]^{-1}$ and (18) coincides with (17).

If $x(t)$ is wide-sense Markov, the multivariate prediction problem is already solved by (14) and (15). For a prediction interval $\alpha$, the best (minimum-mean-square) linear prediction of $x(t+\alpha)$ based on the present and past of $x(t)$ is

$$\hat{E}[x(t+\alpha)\,|\,x(s), s < t] = \hat{E}[x(t+\alpha)\,|\,x(t)] = R(t+\alpha,t)x(t). \tag{19}$$

[3] A covariance matrix $P(t,t)$ is singular if and only if the components $x_k(t)$ of the vector $x(t)$ are dependent in the sense that there exists a set of (complex) scalars $a_k$ such that $E|\sum_k a_k x_k(t)|^2 = 0$ and not all $a_k = 0$. This is proved by noting that the determinant of the matrix $P(t,t)$ is a Gram determinant.

While the wide-sense Markov property is usually not directly verifiable, there is a simple criterion which identifies such processes. This test is a result of

Theorem 2: $x(t)$ is wide-sense Markov if and only if, for all $s \leq t \leq u$,

$$R(u,s) = R(u,t)R(t,s). \tag{20}$$

Proof: In (13), make the identification $t_n = u$, $t_{n-1} = t$, and $t_{n-2} = s$. By an argument identical with the one leading to (16), we have

$$P(u,s) - R(u,t)P(t,s) = 0. \tag{21}$$

When (21) is post-multiplied by $C^*E^{-1}C$ and its terms rearranged, (20) results.

To prove the "if" part of the theorem, we show first that

$$\hat{E}[x(t_n)|x(t_1),x(t_2),\ldots,x(t_{n-1})] = R(t_n,t_{n-1})x(t_{n-1}). \tag{22}$$

Indeed, (22) is equivalent to $P(t_n,t_k)-R(t_n,t_{n-1})P(t_{n-1},t_k) = 0$, $k=1,2,\ldots,n-1$, from (10). But orthogonality is implied by

$$P(t_n,t_k) - R(t_n,t_{n-1})P(t_{n-1},t_k) = 0, \tag{23}$$

which follows from (20) upon taking $t_k = s$, $t_{n-1} = t$, and $t_n = u$.

Equation (22) having been shown, we need only make the further remark that

$$\hat{E}[x(t_n)|x(t_{n-1})] = R(t_n,t_{n-1})x(t_{n-1}) \tag{24}$$

irrespective of the wide-sense Markov property. Together, (22) and (24) imply (13).

As a specialization of the foregoing, we discuss wide-sense Markov processes which are also stationary in the wide sense. These have found some applications in the physics literature, where the emphasis is on gaussian processes (which are then also stationary and Markovian in the usual sense) [6].

We shall define a multivariate wide-sense stationary process as being characterized by the relation

$$P(s,t) = P(s-t) , \tag{25}$$

with the additional proviso that P(0) is nonsingular. From its general definition, R will be seen to assume the form

$$R(t) = P(t)[P(0)]^{-1} . \tag{26}$$

When a process is both wide-sense Markov and stationary, R satisfies both (26) and (20), so that

$$R(s+t) = R(s)R(t) . \tag{27}$$

In fact, the form can now be specified precisely. We have

<u>Theorem 3</u>: Let $x(t)$ be wide-sense stationary and wide-sense Markov. If $R(t)$ is continuous at $t = 0$ (from the right),

$$R(t) = e^{Ct} \qquad t > 0 , \tag{28}$$

where C is a constant matrix all of whose eigenvalues have negative real parts. Conversely, if $R(t)$ satisfies (28), $x(t)$ is wide-sense Markov.

<u>Remark</u>: For $t < 0$, $R(t) = e^{C^*|t|}$.

<u>Proof</u>: That (27) implies (28) follows for matrices just as it does for functions; the uniqueness of (28) follows from the right continuity of $R(t)$ at the origin.[4] For a complete proof, including a verification that the eigenvalues of C have negative real parts, see [5], Theorem 3. To prove the converse part of the theorem, one simply notes that the $R(t)$ given by (28) satisfies (27), so that $x(t)$ is indeed wide-sense Markov.

In making an optimum estimate of quantities such as

$$z(t) = \int_{-\infty}^{\infty} W(t,s)x(s)\,ds \tag{29}$$

[4]It is known that any solutions of (27) not continuous everywhere must be unbounded in every interval.

with x(t) wide-sense Markov, additional difficulties appear. If, however, $\hat{E}$ and integration can be interchanged,

$$\hat{E}[z(t)\,|x(t)\epsilon M] = \int_{-\infty}^{\infty} W(t,s)\,\hat{E}[x(s)\,|x(s)\epsilon M]ds \quad , \tag{30}$$

so that it may be possible to calculate the optimum estimate quite easily.

The following theorem gives conditions under which (30) is valid:

Theorem 4: The interchange of $\hat{E}$ and integration, (30), is valid if, for each t,

$$\int_{-\infty}^{\infty} \|W(t,s)x(s)\|ds < \infty \ . \tag{31}$$

The proof of Theorem 4 requires a working knowledge of measure theory and functional analysis, and appears in [5].

Using (30), it is possible to obtain an explicit expression for the optimum estimate of z(t), based on the entire past of x(t). The integral on the right side of (30) is split into the intervals $-\infty$ to t and t to $\infty$; in the first interval, $\hat{E}[x(s)\,|x(\tau), -\infty < \tau \leq t] = x(s)$, and in the second, (14) applies. Then

$$\hat{E}[z(t)\,|x(\tau), \tau \leq t] = \int_{-\infty}^{t} W(t,s)x(s)\,ds + \left\{\int_{t}^{\infty} W(t,s)R(s,t)\,dt\right\} x(t) . \tag{32}$$

Optimum finite memory filters are also of interest. In terms of the set A introduced earlier [see (14)], a finite memory filter with memory T corresponds to an interval A running from t-T to t. To operate with finite memory filters, it is necessary to use the following property of wide-sense Markov processes: if $\inf_{s\epsilon A} s = s_* \geq t$, then

$$\hat{E}[x(t)\,|x(s), s\epsilon A] = \hat{E}[x(t)\,|x(s_*)]. \tag{33}$$

Thus, the best estimate of x(t) for any time more remote than the length of the filter memory is based on $x(s_*)$, where $s_*$ is the earliest time in the filter memory.

That (33) is true is a consequence of

<u>Theorem 5</u>: If $x(t)$ is wide-sense Markov,

$$R(s,u) = R(s,t)R(t,u) \qquad s \leq t \leq u . \tag{34}$$

Moreover, $x(t)$ is wide-sense Markov if and only if $x(-t)$ is wide-sense Markov.

<u>Proof</u>: For convenience, we write $[P(t,t)]^{-1}$ even if $P(t,t)$ has no inverse; in the latter case, $[P(t,t)]^{-1}$ is to be interpreted as $C^*E^{-1}C$ (where C and E are understood to depend on the parameter t). Suppose $x(t)$ to be wide-sense Markov. Then (20) gives

$$R^*(u,s) = R^*(u,t)R^*(t,s) \qquad s \leq t \leq u . \tag{35}$$

But $P^*(s,t) = P(t,s)$, so that $R^*(s,t) = [P(t,t)]^{-1}P(t,s)$. Now apply this result to (35), premultiply both sides by $P(s,s)$, and postmultiply by $[P(u,u)]^{-1}$. The final outcome of these operations is (34).

To prove the second assertion of the theorem, let $\tilde{x}(t) = x(-t)$, and denote the normalized covariance matrix of $\tilde{x}(t)$ by $\tilde{R}(s,t)$. It is then easily verified that $\tilde{R}(s,t) = R(-s,-t)$. Now choose $-u \leq -t \leq -s$, yielding

$$R(-u,-s) = R(-u,-t)R(-t,-s) \tag{36}$$

from (34). If (36) is rewritten in terms of R, we see that $\tilde{R}(u,s) = \tilde{R}(u,t)\tilde{R}(t,s)$ with $s \leq t \leq u$ (which follows from $-u \leq -t \leq -s$). Thus $\tilde{x}(t)$ is wide-sense Markov.

The "if" part of the theorem follows immediately. For, if $\tilde{x}(t)$ is wide-sense Markov, so is (by what we have just proved) $\tilde{x}(-t) = x(t)$.

With the aid of Theorem 5, it is easy to verify (33). Let $-A$ be the set containing $-t$ if and only if $t \epsilon A$. We thus have, taking $(-s)^* = \sup_{-s \epsilon -A} (-s) \leq -t$,

$$E[\tilde{x}(-t)\,|\,x(-s),(-s)\epsilon -A] = E[\tilde{x}(-t)\,|\,x((-s)^*)] . \tag{37}$$

With the change from $\breve{x}$ to x, (37) becomes

$$E[x(t)\,|\,x(s), s\epsilon A] = E[x(t)\,|\,x(-(-s)^*)] \,. \tag{38}$$

We complete the calculation by showing that $-(-s)^* = s^* \geq t$. Indeed,

$$-(-s)^* = -\sup_{s\epsilon A}(-s) \geq t, \text{ and } -\sup_{s\epsilon A}(-s) = \inf_{s\epsilon A} s = s_* \,.$$

## III. APPLICATIONS

With the theoretical background that has been developed, it is possible to solve prediction and estimation problems not amenable to the usual techniques. Solutions to several apparently difficult problems will be exhibited in closed form without necessarily exhausting the power of wide-sense Markov theory.

Consider again the problem of estimating a z(t) defined by (29), W(t,s) representing some desired operation on x(t). It is assumed that x(t) is available over a finite segment of the past, so that the optimum finite memory filter is sought. To compute $\hat{E}[z(t)\,|\,x(\tau), t-T \leq \tau \leq t]$, $\hat{E}$ and the integration are interchanged, and the integral is split into the three intervals $-\infty$ to t-T, t-T to t, and t to $\infty$. Equation (33) may be applied to the first integral, and (14) to the third. In the second integral, advantage is taken of the fact that $\hat{E}[x(s)\,|\,x(\tau), t-T \leq \tau \leq t] = x(s)$ whenever $t-T \leq s \leq t$. The result of the computation is therefore

$$\begin{aligned}\hat{E}[z(t)\,|\,x(\tau), t-T \leq \tau \leq t] &= \left[\int_{-\infty}^{t-T} W(t,s)R(s,t-T)\,ds\right] x(t-T) \\ &\quad + \int_{t-T}^{t} W(t,s)x(s)\,ds + \left[\int_{t}^{\infty} W(t,s)R(s,t)\,ds\right] x(t)\,.\end{aligned} \tag{39}$$

The finite memory filter which operates on x(t) to yield $\hat{E}[z(t)\,|\,x(\tau), t-T \leq \tau \leq t]$ can be described by a weighting function G(t,s) such that

$$\hat{E}[z(t)\,|\,x(\tau), t-T \leq \tau \leq t] = \int_{t-T}^{t} G(t,s)x(s)\,ds \,. \tag{40}$$

It is convenient to write $G(t,s)$ in terms of the $\delta$-function and the unit step function $\lambda(t)$; the latter is zero for negative argument and unity for positive argument. Both step and $\delta$-functions are regarded as scalar multipliers. The equation for $G(t,s)$ reads:

$$G(t,s) = \left[\int_{-\infty}^{t-T} W(t,u)R(u,t-T)du\right]\delta(t-T-s) + W(t,s)[\lambda(s-t-T)-\lambda(s-t)]$$
$$+ \left[\int_{t}^{\infty} W(t,u)R(u,t)du\right]\delta(t-s) . \qquad (41)$$

If $W(t,s)$ is zero for $s \leq t$, it is called wholly unrealizable. Then only the final term of (39) or (41) remains, and (independently of $T$) the estimate of $\hat{E}[z(t)|x(\tau),\ t-T \leq \tau \leq t]$ is obtained by applying $n^2$ time-varying multipliers and $n$ summers to $x(t)$. If, in additon, $W(t,s)$ is time invariant, i.e., $W(t,s) = W(t-s)$, and $x(t)$ is wide-sense stationary (as well as wide-sense Markov), the $n^2$ multipliers have constant gains.[5] In particular, the jth component of $\hat{E}[z(t)|x(\tau),\ t-T \leq \tau \leq t]$ is $\sum_{1}^{n} g_{jk}x_k(t)$, where $g_{jk}$ is the jk component of the constant matrix

$$G = \int_{-\infty}^{0} W(u)e^{-C^*u}du . \qquad (42)$$

A particular example of a wholly nonrealizable $W$ is that associated with prediction over an interval $\alpha$. There, $W(s,t) = \delta(t+\alpha-s)I$. Substituting this value of $W$ in (41) (only the last term need be considered) gives a result identical with the earlier computation (19).

The general results (39) and (41) lend themselves also to other applications. Assume a sampled data system, in which only one sample is available to the filter memory. If the leading edge of the sample is at time $s_*$, and the trailing edge at time $s^*$, (41) becomes

$$G(t,s) = \left[\int_{-\infty}^{s_*} W(t,u)R(u,s_*)du\right]\delta(s_*-s) + W(t,s)[\lambda(s-s_*)-\lambda(s-s^*)]$$
$$+ \left[\int_{s^*}^{\infty} W(t,u)R(u,s^*)ds\right]\delta(s^*-s) . \qquad (43)$$

[5]Apart from variations in the signs of the multiplier gains, n operational amplifiers with adjustable input resistors suffice to mechanize the optimum system.

Special cases are readily deduced from (43). It is clear, for instance, that only the trailing edge of a sample (the trailing edge of the last sample if there are more than one) is employed if $W(t,s)$ is wholly unrealizable; this trailing edge is passed through a set of time-varying amplifiers and summed. Another specialization results if the sample is instantaneous (zero width). Then $s_* = s^*$, the second term of (43) disappears, and there remains

$$G(t,s) = \left[\int_{-\infty}^{\infty} W(t,u)R(u,s^*)\,du\right] \delta(s-s^*) \ , \tag{44}$$

which is to say,

$$\hat{E}[z(t)\,|x(s^*)] = \left[\int_{-\infty}^{\infty} W(t,u)R(u,s^*)\,du\right] x(s^*) \ . \tag{45}$$

So far, there has been no indication when wide-sense Markov processes might be encountered. In fact, such processes do constitute the output of at least one important class of linear systems. Considered will be systems described by linear time-varying differential equations with white-noise forcing.[6] For these systems, $R(s,t)$ will be computed explicitly, thus solving the prediction problem completely for outputs of these systems.

For a first-order equation,

$$\dot{x} + f(t)x = h(t)n(t) \qquad x(0) = 0 \ , \tag{46}$$

in which all quantities are real scalars (matrices with only one entry), and $n(t)$ is a (real) random process with covariance $E[n(s)n(t)] = \delta(s-t)$.

Evidently, $x(t)$ is the output of a time-varying system whose input consists of white noise modulated by $h(t)$. For instance, $h(t)$ becomes a pulse train if the system receives its input through a (possibly non-periodic) sampling switch. More generally, $h(t)$ reflects the variation in noise strength with time.

If one takes

$$H(s,t) = \exp\left\{+\left[\int_{s}^{t} f(u)\,du\right]\right\} , \tag{47}$$

[6] Compare [3].

the solution of (46) becomes

$$x(t) = \int_0^t H(t,u)h(u)n(u)du = H(t,0)\int_0^t H(0,u)h(u)n(u)du \qquad (48)$$

since $H(s,u) = H(s,t)H(t,u)$. It is now easy to compute $R(s,t) = P(s,t)[P(t,t)]^{-1}$. In this case, the inverse becomes the reciprocal, so that we have

$$R(s,t) = \frac{E[x(s)x(t)]}{E[x(t)^2]} . \qquad (49)$$

An easy computation yields

$$E[x(s)x(t)] = H(s,0)H(t,0)\int_0^{\min(s,t)} H^2(0,u)h^2(u)du , \qquad (50)$$

which we rewrite as

$$E[x(s)x(t)] = H(t,0)\int_0^{\min(s,t)} H(s,u)H(0,u)h^2(u)du . \qquad (51)$$

From (51) and (49) we deduce that

$$R(s,t) = \frac{\int_0^s H(s,u)H(0,u)h^2(u)du}{\int_0^t H(t,u)H(0,u)h^2(u)du} \qquad (52)$$

whenever $s \leq t$. The expression for $R(s,t)$ with $s \geq t$ is even simpler. We now use (50), and note that the integral portion of this equation remains unchanged if we set $s = t$. Therefore, the integral term cancels when we compute $R(s,t)$. Using the relation $H(u,v) = 1/H(v,u)$, we have for our result

$$R(s,t) = H(s,t) \qquad s \geq t . \qquad (53)$$

The Markov property is now easily verified by substituting (53) into one side of (20) and using the identity $H(s,u) = H(s,t)H(t,u)$. Verification may also be accomplished through use of (52), if desired.

It is of course possible that the denominator of (52) is zero; this occurs only when $h(u) = 0$ in the interval zero to t. In such an event, we take $R(s,t) = 0$, which also makes the estimate of any $z(t)$, based on $x(\tau)$, $0 \leq \tau \leq t$, be zero. This is to be expected, since $x(t) = 0$ over the same interval.

Since $R(s,t)$ may be directly computed from (52) or (53), any of the formulas of this section may be applied. For instance, the optimum prediction over interval $\alpha$ is

$$\hat{E}[x(t+\alpha)\,|\,x(\tau), 0 \leq \tau \leq t] = H(t+\alpha,t)x(t), \tag{54}$$

which shows that prediction may be accomplished by a time-varying multiplier. When $W(t,s)$ is wholly nonrealizable, other optimizations are comparably simple.[7]

The above calculations may be extended to matrix differential equations, which are written

$$\dot{x} = A(t)x + M(t)n(t) \qquad x(0) = 0, \tag{55}$$

all symbols representing matrices, which may be complex. The "white noise" is now characterized by $E[n(s)n^*(t)] = \delta(t-s)I$, where I is the identity matrix.

The solutions to optimization problems involving the $x(t)$ given by (55) are formulated in terms of the fundamental matrix, $X(t)$, which provides the solution to

$$\dot{X} = A(t)X, \qquad X(0) = 0. \tag{56}$$

That a knowledge of $X(t)$ is required implies no loss in generality, since $X(t)$ is required even to compute $P(s,t)$, or to solve (55) with non-stochastic forcing.

The calculations are similar to those made for the first-order equation, except that matrices fail to commute, and reciprocals of singular matrices cannot be identified with inverses. It remains true, however, that there is an $H(t,s)$ such that

[7]For a more complete discussion of the optimum estimator problem as applied to (46), see [7].

$$x(t) = \int_0^t H(t,u)\,M(u)\,n(u)\,du \tag{57}$$

with

$$H(s,t) = X(s)\,[X(t)]^{-1} ,$$

a nonsingular matrix having the property $H(s,t) = [H(t,s)]^{-1}$.

Proceeding as before,

$$P(s,t) = \int_0^{\min(s,t)} H(s,u)\,M(u)\,M^*(u)\,H^*(t,u)\,du . \tag{58}$$

For ease of computation, it is sometimes advantageous to write (58) as

$$P(s,t) = X(s)\left\{\int_0^{\min(s,t)} [X(u)]^{-1}M(u)\,M^*(u)\,[X^*(u)]^{-1}du\right\}X^*(t) \tag{59}$$

If the integral

$$V(t) = \int_0^t [X(u)]^{-1}M(u)\,M^*(u)\,[X^*(u)]^{-1}du \tag{60}$$

is invertible, we also have

$$[P(t,t)]^{-1} = [X^*(t)]^{-1}[V(t)]^{-1}[X(t)]^{-1} , \tag{61}$$

so that

$$R(s,t) = H(s,t) \qquad s \geq t \tag{62}$$

and, for $s \leq t$,

$$R(s,t) = X(s)\,V(s)\,[V(t)]^{-1}[X(t)]^{-1} = X(s)\,V(s)\,[X(t)\,V(t)]^{-1} . \tag{63}$$

Either (62) or (63) may be used to verify that $x(t)$ is wide-sense Markov. This is easily accomplished by substitution of (62) or (63) into (20).

It may be shown (see [5]) that (60) is nonsingular if $M(u)$ is nonsingular in some neighborhood of the origin, so that a simple sufficiency condition checks the applicability of the above formulas. On the other hand, there are some applications in which (60) is clearly singular. For instance, we may choose $M(u)$ to possess only one nonzero element, corresponding to white noise imposed on only one of the vector components. One such form of $M(u)$ leads to the scalar equation.

$$a_n(t)\,\frac{d^n x}{dt^n} + a_{n-1}(t)\,\frac{d^{n-1}x}{dt^{n-1}} + \cdots + a_o(t)x \;=\; n(t) \;,$$

which is, of course, covered by our theory.

The case of a singular $V(t)$ again requires us to find eigenvalues and eigenvectors. Since $V(t)$ is Hermitian, we may define $C_t$, $D_t$, and $E_t$ to have the same meaning as in Section II, except that they apply to the matrix $V(t)$. It is then possible to show (see [5]) that

$$R(s,t) \;=\; X(s)\,[C_s D_s E_s^{-1} C_s]\,[X(t)]^{-1} \qquad s \geq t \;. \tag{64}$$

The $R(s,t)$ of this form also satisfies (20), so that $x(t)$ is wide-sense Markov in any case (see [5] for proof). If $V(t)$ is nonsingular, (64) reduces to (62).

Since the determination of $R(s,t)$ shows $x(t)$ to be wide-sense Markov, substitution into any of the earlier optimization formulas is appropriate. For an infinite memory filter, and/or a $z(t)$ generated by a wholly unrealizable weighting function, only the simple expression (62) appears in the result. The prediction formula, for example, is precisely (54).

## REFERENCES

[1] N. Wiener and P. Masani, "Multivariate Stochastic Processes," Acta Math., 98, 111-150, 1957; 99, 93-137, 1958.

[2] R. Amara, "Linear Least Square Synthesis of Multivariate Control Systems," AIEE Transaction Paper 59-214, February 1959.

[3] R. Kalman and R. Bucy, "New Results in Linear Filtering and Prediction Theory," presented at the Joint Automatic Control Conference, Cambridge, Mass., September 1960, as Paper No. 60-JAC-12.

[4] M. Akhiezer and I. Glazman, Theory of Linear Operators in Hilbert Space, translated by M. Nestell, Ungar, New York, 1961, Vol. I.

[5] F. Beutler, "Multivariate Wide-Sense Markov Processes and Prediction Theory", Ann. Math. Stat., in press.

[6] M. Wang and G. Uhlenbeck, "On the Theory of Brownian Motion II," Rev. of Modern Phys., 17, 323-342, 1945.

[7] F. Beutler, "Prediction for Wide-Sense Markov Processes," IRE Trans. Information Th., IT-7, 267-269, 1961.